



Data Opportunities: Building Capacity and Linkages with Data Science Communities

By
Thilanka Munasinghe and Kathy Fontaine
RPI



About Us

- Thilanka Munasinghe is a Lecturer at RPI
 - His focus is on data science and data analytics as applied to UN Sustainable Development Goals (SDGs).
- Kathy Fontaine is a Sr. Research Scientist and Adjunct Professor at RPI, and a member of the GEO Programme Board [representing ESIP].
 - Her focus is the intersection of policy and data science.

Why are we Here?

- RPI's Data Science students and staff are a potentially valuable resource for GEO. Why? You'll see in the upcoming slides.
- Semester-long class projects and Independent Study projects are opportunities for GEO to “get some data science done”
 - Taking an initial look at well-defined questions
 - Potentially validating existing work
 - Applying existing work to new data sets
- How can we help EO4Health?

Some Definitions

- Data Science - Data science is an inter-disciplinary field that uses scientific methods, processes, algorithms in managing a digital data collection, that focus on Data Life-Cycle and Data Management ... etc..
- Data Analytics - Data analytics is the science of analyzing data in order to make meaningful conclusions using statistics, machine learning and other techniques... etc.
- Xinformatics - The science of processing and obtaining information from data and study information systems. X can be “health”, “astro”, etc..
- Co-Terminal Students - undergraduates who are simultaneously taking coursework towards a Master's Degree

What we have done so far:

- During the Data Science, Data Analytics, and Informatics course projects, we ask students to use publicly available datasets from UN, CDC, NOAA, NASA, USDA, etc. to work on their semester-long project.
- During the past semesters, some of the class projects have been published as full-length papers and presented as posters in international conferences such as IEEE BigData, ACM ICMHI (International Conference in Medical and Health Informatics), and AAAI AI for Social Good Symposium.

Introduce projects related to EO4Health

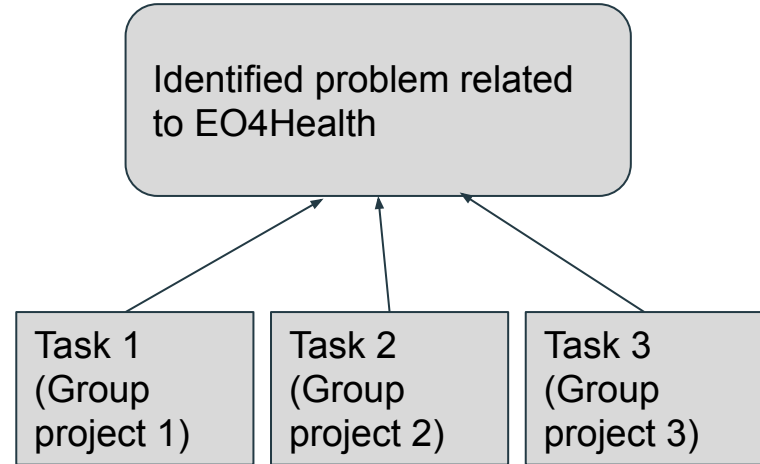
- During the upcoming semesters (Spring 2021 onwards), we are planning to introduce some project ideas focusing on EO4Health to give as class projects to students in the Data Science, Data Analytics and XInformatomics courses at RPI

Semester long class projects

- Data Science and Xinformatics - Group project
- Data Analytics course - Individual Project
- Independent Studies
- Co-Terminal Students

Breaking down a larger project idea to subparts for group projects

- During the group and individual projects, we plan to break down an identified bigger problem into sub-tasks/problems and assign those sub-tasks to groups and ask them to work as their semester project. Later we intend to combine group projects that related to the same problem.



Dissemination of the project results:

- Students are encouraged and asked to submit their project findings as conference papers at the end of the semester to various conferences as did during the past semesters.
- Students are also asked to create a poster and present their findings during the RPI Data Science and Data Analytics Poster Session at the end of the Fall 2021 semester.
- Develop interactive visualizations and *Dashboards* based on the results, may be host them on ITWS and RPI websites.

Next Steps

- Talk in your small groups about potential projects
- Get in touch with Thilanka or Kathy to have a more in-depth conversation about the possibilities, answer questions, identify data sets, and so on.
 - We could speak at your group meetings....
- No time limit on these - there are classes every semester, so projects could start whenever.

Contact Information

- Thilanka Munasignhe
 - munast@rpi.edu
- Kathy Fontaine
 - fontak@rpi.edu

Thank You!

Questions?